DOCUMENT RESUME

ED 124 580 TM 005 335

AUTHOR Hogan, M. J.; And Others
TITLE The Effect of Individual Differences in Abilities on Interpreting the Dimensionality of Multiple-Choice Tests.
SPONS AGENCY National Fund for Medical Education, Cleveland, Ohio.
PUB DATE [Apr 76]
NOTE 4p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; *Aptitude Tests; Discriminant Analysis; *Factor Analysis; Higher Education; *Individual Differences; Medical Students; *Multiple Choice Tests; *Test Interpretation

ABSTRACT

The problem of interpreting the dimensionality underlying binary test data is addressed. Medical school students were tested on both achievement and aptitude measures. The person-by-test (aptitude) interaction matrix was factor-analyzed to identify patterns of individual differences. The resultant groups of similar patterns were used as the criteria in a discriminant function analysis where the discriminator variables are the achievement test items. The results strongly indicate that performance on the achievement test can be partially explained by the differential pattern of performance on the aptitude tests. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED124580

The Effect of Individual Differences in Abilities on Interpreting the Dimensionality of Multiple-Choice Tests[1]

HOGAN, M.J., GALLAGHER, R.E., SIROTKIN, R.A., WOLFE, B.L., and SCALZI, P.M.
Wayne State University, School of Medicine

Ralph Sirotkin
Division of Educational Services and Research
Wayne State University, School of Medicine
540 E. Canfield
Detroit, Michigan 48201

Introduction

Factor analysis has been the predominant technique used to interpret the structure of multiple-choice (M-C) achievement tests with binary response data. The major difficulty with this approach lies in the meaningful interpretation of the resultant factor pattern matrix. This difficulty stems from the fact that the number of factors extracted often does not coincide with the number of factors predicted by theory.

Guilford (1941) and Horst (1966) suggested that the number of factors extracted will be determined by the number of levels of difficulty in the test, even though the items themselves were homogeneous in respect to content McDonald and Ahlawat (1974) asserted that "spurious" factors - such as those correlating with difficulty - are an artificat of the linear factor analytic model. The implication of the McDonald-Ahlawat assertion is that there is no such thing as a "difficulty" factor; rather, that when imposing a linear factor model on curvilinear data, factors interpreted as "spurious" or due to "difficulty" are, in fact, factors due to a curvilinear relation existing between test scores and factor content.

Gollob (1968a, 1968b) suggested that the item and person main effects are confounded with the person by item interaction. In attempting to develop a model for investigating individual differences, he maintained that a meaningful distinction exists between person main effects and patterns of individual differences as represented in the person by test interaction. The person main effect represents the individual's average performance across tests in relation to other individuals. The pattern of the individual's performance, on the other hand, represents how that individual performs on any one test in relation to all other tests in the sample. In effect, this pattern of performance is represented in the person by test interaction term in the ANOVA model. Gollob's model, the FANOVA model, is essentially a factor analysis of this interaction term. This is accomplished by subtracting out both row (person) and column (test) main effects to double center the data matrix, then factoring the residual matrix (person by test interaction).

Overall and Klett (1972) have incorporated Gollob's logic into a Linear Typal Analysis Program that is similar to profile analysis. This program attempts to describe the dimensionality of the person by test interaction component of the total variance from a person, or Q, point of view, thus grouping the patterns of individual differences according to their similarity

This study hypothesized that the patterns of individual differences resulting from student performance on aptitude tests will prove useful in establishing a taxonomy of M-C items that will facilitate the interpretation of the dimensionality of the M-C test. The connection between the patterns of individual differences and M-C items were established through the use of a multiple discriminant function technique, where the M-C items were the discriminator variables, and the groups of patterns of individual differences were the criteria.

[1]This project was supported in part by a grant from The National Fund for Medical Education.

TM005 335

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
M.J. HOGAN
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Method

The 112 subjects used in the analysis were drawn from 256 first year students at Wayne State University School of Medicine.

The M-C items were selected from an examination covering the introductory unit composed of biochemistry, gross anatomy and cell biology. A set of thirty-five (35) items were chosen according to two criteria; (1) a point bisereal coefficient between item scores and test scores of greater than .15 and, (2) an item difficulty of greater than .15 and less than .85. This set of thirty-five (35) items was used as discriminator variables in the discriminant function analysis.

The aptitude measures were based on the Structure of Intellect (SI) model of Guilford (1974). The SI tests used in this study were chosen on the basis of an *a priori* analysis of the relationship between the items on multiple choice examinations from previous years and the abilities the items were purported to measure. As a result of this analysis, tests representing eighteen (18) cells of the Guilford cube were selected with two (2) tests per cell.

The achievement tests were administered in the normal course of the term. The aptitude tests were administered to the entire first year class as a group in one day, prior to beginning their academic studies.

Results

The principal axes method of factor analysis with iterations for communality estimates was used in an initial attempt to understand the dimensionality underlying the M-C test. The factors were rotated to a simple structure criterion using the direct oblimin method of rotation. This method could yield correlated factors. However, in all analyses, the correlations among factors was low.

The factor analysis of the thirty-five (35) items was not unequivocally interpretable by any frame of reference. Attempts were made to interpret this matrix using content, item difficulty, and a revised Bloom taxonomy with no success.

Six of the SI tests were excluded from the analysis due to extreme skewness. The Linear Typal Analysis of the thirty (30) remaining tests indicated three clear groups of patterns of performance. The three (3) groups were identified as evaluation, cognition, and memory.

The multiple discriminant function analysis resulted in a significant set of discriminant functions ($F=1.70$, $df=40, 180$, $p<.05$) and included twenty (20) of the thirty-five (35) items entered. It was quite apparent which items were the best discriminators between the groups.

The factor pattern matrix of the thirty-five (35) M-C items was re-interpreted on the basis of the discriminant function analysis from the point of view of the relationship between that item and the group for which it best discriminated. That is to say, if the item discriminated the cognition group from the evaluation and memory groups, it was classified as a cognition question. If an item discriminated the evaluation group from the cognition and memory groups, it was classified as an evaluation question. And, if an item discriminated the memory group from the evaluation and cognition groups, it was considered a memory question. This re-interpretation of the factor pattern matrix, using the new taxonomic scheme, made it quite apparent that the dimensions underlying the multiple choice test were, to a large extent, representative of the type of ability called for to answer the question.

Discussion

One of the major findings of this study was that the interpretation of the dimensionality of an M-C examination from an individual differences point of view was useful towards understanding that dimensionality. However, the fact that this interpretation was not unequivocal stresses the confounding of the item main effect with the person by item interaction.

The results reported in this study raise several questions regarding the interpretation of M-C examinations. First, since it has been shown that taxonomic systems can be further developed to explain the dimensionality of an examination, is it fruitful to continue using total scores with the concomitant loss of information in assessing performance?

Second, it is readily apparent that persons differ in regard to ability, and M-C items are differentially related to these abilities, therefore, are achievement examinations being constructed that are biased against groups of students due to over-emphasis on the way in which items are constructed?

Third, when there is a strong person by test interaction that is not taken into account, might not the interpretation of the dimensionality of the M-C items be misleading and/or spurious?

REFERENCES

Carroll, J.B. The effect of difficulty and chance success on correlations between items or between tests. Psychometrika, 10, 1-19, 1945.

Gollob, H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika, 33, 73-115, 1968a.

_____. Confounding of sources of variation in factor-analytic techniques. Psychological Bulletin, 70, 330-344, 1968b.

Guilford, J.P. The difficulty of a test and its factor composition. Psychometrika, 6, 67-77, 1941.

_____, and Hoepfner, R. The Analysis of Intelligence. New York: McGraw-Hill, 1971.

Horst, Paul. Psychological Measurement and Prediction. Belmont, California: Wadsworth, 1966.

McDonald, R.P. and Ahlawat, K.S. Difficulty factors in binary data. Br. J. Math. Statist. Psychol., 27, 82-99, 1974.

Overall, J.E. and Klett, C.J. Applied Multivariate Analysis. New York: McGraw-Hill, 1972.